How to localize 3D-views in space?

Nikolaus Schön, Gerd Häusler

Max Planck Research Group, Institute of Optics, Information and Photonics, University of Erlangen-Nuremberg

mailto:nikolaus.schoen@optik.uni-erlangen.de

We describe an algorithm to automatically align 3D views of an object (*registration*), acquired by an optical 3D sensor from different directions. The method is based on the selection and assignment of "salient" points, with respect to information theoretical considerations. The approach is implemented and we demonstrate its ability to reliably register 3D-views of a wide range of objects, within a few seconds.

1 Introduction

A variety of applications use three dimensional models, e. g. CAM-based reproductions of works of art or dental inlays, status control in facial surgery, or model based image analysis [1].

Complete 3D models are reconstructed from many single 3D views, which must first be aligned to each other. This process is called registration and is generally split up in two steps: coarse registration and fine registration. Fine registration presupposes that two given data sets are already approximately aligned and it minimizes iteratively the residual sum of distances between closest points on two surfaces. For this problem, very efficient optimization methods are known [2], and a theory exists to predict the achievable accuracy of fine registration [3].

For coarse registration, however, many surface reconstruction approaches rely on manual interaction: A human selects three pairs of corresponding salient points on the surfaces. From these correspondences, the transformation parameters for an initial coarse alignment are computed.

The automatic method of Winkelbach et al. [4] relies mainly on exploiting geometrical constraints, that are determined by the shape of the data sets. Local surface features (curvatures) are used only supplementally. In contrast, the most important step in our method is the selection of points with highly salient and discriminant features.

2 Selection of Salient Points

It depends on the neighborhood of a point whether it is a salient point. On a sphere, for example, we can not localize a point, because spheres are completely symmetric surfaces. Therefore, points with a spherical neighbourhood should be regarded as non-salient, whereas points with a strongly structured neighborhood should get a high salience value.

For each point $\mathbf{p}$ on the surface, a salience value is computed, that reflects the content of information in a small neighborhood around $\mathbf{p}$. A local coordinate system $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ is defined (see Fig. 1), with origin $\mathbf{p}$, where $\mathbf{e}_z$ is identical with the normal $\mathbf{n}$ in $\mathbf{p}$.

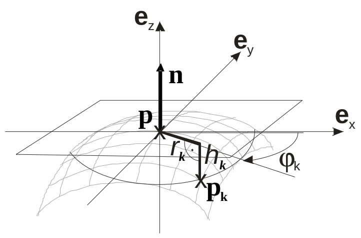

Fig. 1 *Local coordinate system in* $\mathbf{p}$.

For each point $\mathbf{p}_k$ the neighborhood, with cylindrical coordinates (r_k, φ_k, h_k), we define a virtual sphere that contains $\mathbf{p}$ and $\mathbf{p}_k$ and has in $\mathbf{p}$ the normal $\mathbf{n}$. The curvature κ_k of the sphere is

$$\kappa_k = \frac{2h_k}{{r_k}^2 + {h_k}^2}. \tag{1}$$

Generally, each point $\mathbf{p}_k$ defines another sphere with another curvature κ_k. If, however, the neighborhood of $\mathbf{p}$ is spherical, we get the same κ-value for all $\mathbf{p}_k$. An appropriate salience measure is therefore the entropy of the local κ-values. We quantize κ to N discrete values $\hat{\kappa}_i$, $i = 1, \ldots, N$, and estimate the relative frequencies $p(\hat{\kappa}_i)$ of their appearance by computing a histogram. The information entropy of κ, estimated by

$$H_{\mathbf{p}} = -\sum_{i=1}^{N} p(\hat{\kappa}_i)\log_2(\hat{\kappa}_i) \qquad (2)$$

actually is a measure of point salience: For spherical and planar neighborhoods, it has the value $H_{\mathbf{p}} = 0$, whereas for neighborhoods with strongly varying κ-values, $H_{\mathbf{p}}$ is big. From the data set, those points are selected for coarse registration, where

1) $H_{\mathbf{p}}$ has a high value and

2) $H_{\mathbf{p}}$ is locally maximal, i.e. higher than the $H_{\mathbf{p}'}$-value of any direct neighbor point $\mathbf{p}'$.

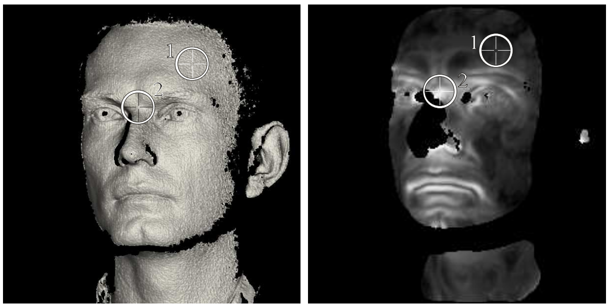

Fig. 2 *Range image of a face (top) and salience map; saliences are displayed as intensity values. Two points are emphasized by circles: (1) point with nearly spherical neighborhood, (2) point with highly structured neighborhood.*

Fig. 3 *Visualization of a complete 3D model of a work of art (Horses of Franz Marc, courtesy of "Staatliche Galerie Moritzburg").*

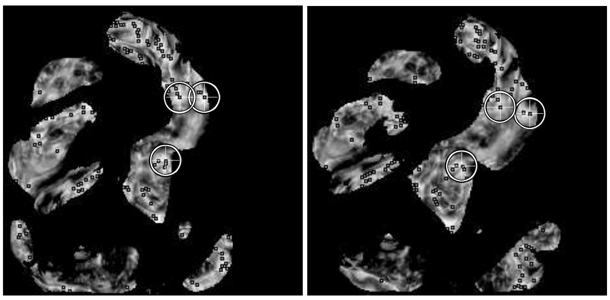

Fig. 4 *Salience maps of two views of the work of art in Fig. 3. The local salience maxima are marked by black squares, three point correspondences and the (approximate) point neighborhoods are emphasized by white circles.*

This information-based point selection (see Fig. 4) is the most essential step in coarse registration, because it reduces the problem to the assignment of a small number of possible point correspondences.

3 Matching

After selection of salient points, their pairwise correspondence is still unclear. Therefore, we encode the neighborhood of each salient point $\mathbf{p}$ as a feature vector, that is attributed to $\mathbf{p}$. Possible point correspondences are established by comparing these feature vectors, using euclidean distance. The feature vector is computed from the set of curvatures κ_k of all neighbor points and the geometric invariants $r_k' = r_k^2$. For the local set of all pairs (r_k', κ_k), a two dimensional histogram is computed, and the bin values of the histogram are used as components of the feature vector. Since there still remain ambiguities in point assignment, we additionally exploit the fact, that the pairwise distances in a group of assigned points have to be consistent over different views.

The complete registration of two 3D views takes only ~20s. Applications of the method to medical and dental data are described in [5], together with further references.

Acknowledgement

This project is funded as subproject A1 of "Sonderforschungsbereich 603", "Model based Analysis and Synthesis of Complex Scenes and Sensor Data" by DFG (German Research Foundation).

References

[1] V. Blanz, Vetter, T., "Face Recognition based on fitting a 3D morphable model", PAMI **25**(9), 1063-1074 (2003).

[2] N. Diehl, H. Burkhardt: "Motion Estimation in Image Sequences" in *International Workshop on High Precision Navigation,* K. Linkwitz and U. Hangleiter, (editors), Springer (1989).

[3] X. Laboureux and G. Häusler, "Localization and Registration of Three-Dimensional Objects in Space - Where are the Limits?" Appl. Optics **40**(29): 5206-5216 (2001).

[4] S. Winkelbach, M. Rilk, C. Schönfelder, F. M. Wahl, "Fast Random Sample Matching of 3d Fragments" in Pattern Recognition (DAGM 2004), Lecture Notes in Computer Science 3175, Springer: 129-136 (2004).

[5] N. Schön, M. Benz, T. Maier, E. Nkenke, F. W. Neukam, G. Häusler, „Informationsoptimierte Merkmale zur Grobregistrierung von Freiform-Flächen" in Bildverarbeitung für die Medizin 2004, Springer (2004).